Optimal Filtering of a Gaussian Signal in the Presence of Lévy Noise

1995

Hyungsok Ahn †‡ and Raisa E. Feldman †
Department of Statistics and Applied Probability
University of California, Santa Barbara

Abstract. Many engineering applications require extracting a signal from observations corrupted by additive noise, possibly heavy-tailed. We assume that the observation noise is a Lévy process, while the signal is Gaussian, and derive a non-linear recursive filter that minimizes the L^2 error. A sub-optimal filter is proposed for numerical purposes, and simulations show that it out-performs the existing linear filter.

Key Words: Nonlinear filtering; Lévy process; infinitely divisible law; non-Gaussian noise.

AMS subject classifications: Primary 60G35, 60H30, 60E07; Secondary 60J30, 60J60.

† This research was partially supported by ONR Grant No. N00014-94-1-0191.
‡ This research was supported by ONR Grant No. N00014-96-1062.

1 Introduction

Estimation problems involving non-Gaussian noise have received much attention in recent years; see for example Kassam (1988) and Nikias and Shao (1995) and references therein. In this paper we consider filtering of Gaussian signals contaminated by independent infinitely divisible noise, a class that includes compound Poisson, alpha-stable, and Gaussian noises as special cases. As in Nikias and Shao (1995), our intent here is to develop a methodology that works in the presence of heavy-tailed, possibly infinite variance, noise.

A practical application in which the observation noise can be represented as a combination of a Gaussian noise and a discrete shock is the tracking error in code tracking loops, e.g., such as those used in Global Positioning System (GPS) receivers. The tracking error, defined as the time offset between the received signal (plus noise) and the internally-generated version of the same signal, would be purely Gaussian except that occasionally the tracking slips forward or backward by one time step, introducing a discontinuity.

Another example can be found in many coherent communication systems, where the receiver must estimate the phase of a *carrier*, a process known as *phase tracking*; see Poor (1988), p524. It is customary to model the unknown phase as a Gaussian process, and in some applications, it is also appropriate to model the noise and interference as Gaussian. When, however, the number of significant interference sources is small and when interferers are pulsed or hopped, the noise and interference may be described better by the combination of a Wiener process and a jump process.

Filtering problems involving heavy-tailed noise were initiated by Stuck (1978). He examines a classic set-up for Kalman filtering with i.i.d. plant and observation noises, each coming from the same symmetric alpha-stable distribution. A linear recursive filter was chosen to satisfy the criterion of minimum dispersion of the prediction error. Cambanis and Miller (1981) consider among other linear problems the linear filtering of a signal in noise when both signal and noise belong to a special class of processes, e.g., harmonizable. Their solution is rather abstract as it is written in terms of integrals with respect to stable measures. Nikias and Shao (1995) consider adaptive Wiener-type filters constructed according to the minimum dispersion criterion. The closest to our work is the paper by Le Breton and Musiela (1993). They derive a linear filter for a continuous time system in which both the signal and observation processes are given by linear equations with respect to general semimartingale noises and show that their filter minimizes L^p error.

Here we adopt a specific class of noise processes and exploit analytic properties of these processes. More specifically, we assume that the *signal* process X and the *observation* process Y evolve according to the following stochastic differential equations:

$$X_t = X_0 + \int_0^t a(s)X_s ds + \int_0^t b(s)dB_s, \tag{1}$$

$$Y_t = Y_0 + \int_0^t c(s)X_s ds + \int_0^t d(s)dW_s + J_t, \tag{2}$$

where B and W are independent standard Brownian motions, and J is a quadratic pure jump Lévy process, that is, a process with stationary independent increments and with

quadratic variation that has no continuous part. See Protter (1990) for details; in particular, such processes are independent of a Brownian motion. The coefficients a, b, c, d are continuous real valued functions of time. We further assume for convenience that X_0 and Y_0 are constants.

If one is allowed to monitor Y continuously in time, one knows exactly when each jump occurs and can extract Y^c, the continuous part of the observation:

$$Y_t^c = Y_0 + \int_0^t c(s)X_s ds + \int_0^t d(s)dW_t. \tag{3}$$

Thus, to estimate the signal X, one would apply the Kalman-Bucy filter to the Y^c extracted from the observation. However, a more realistic assumption would be that the observations arrive in discrete times $0 < t_1 < t_2 < \ldots$, because any digital signal processing system can only sample at a finite rate limited by the speed of an analog-to-digital (A/D) convertor, and can only process observations at a finite rate determined by the available processor power. In this case the jump can not be removed and a different filter is required.

This paper establishes the *optimal* filter based on observations sampled in discrete times, where the optimality refers to minimizing the L^2 error, that is L^2 distance between the signal and the filter output. In other words, we find

$$\hat{X}_{t_j} = E\left[X_{t_j} \middle| \mathcal{F}_{t_j}\right] \tag{4}$$

where $\mathcal{F}_{t_j} = \sigma(Y_{t_1}, \ldots, Y_{t_j}\}$ is the sigma field generated by observations obtained up to time t_j. Since our signal is Gaussian, the integrability assumption is satisfied. Note that we do not restrict our filter to be linear. In fact when the noise is strictly non-Gaussian, i.e., jump noise J is present, (4) yields a non-linear filter. Although a linear filter is more tractable, it could produce a poor estimate in the presence of heavy tailed noise. For example, when J is alpha-stable, the observation process has an infinite variance. Thus, any linear filter would have an infinite L^2 error.

In the remainder of this introduction, we summarize the results in this paper. For simplicity, Δ denotes an increment of a process; for example, $\Delta Y_{t_{j+1}} = Y_{t_{j+1}} - Y_{t_j}$. To obtain the optimal filter, we solve a two stage filtering problem. In the first stage, we compute a *pseudo filter* $\bar{X}$ by assuming that we are able to recognize the proportion of noise contributed by J. That is, $\bar{X}_{t_j} = E\left[X_{t_j} \middle| \mathcal{G}_{t_j}\right]$, where $\mathcal{G}_{t_j}$ is a pseudo filtration that contains the information of $(J_{t_1}, \ldots, J_{t_j})$ as well as $\mathcal{F}_{t_j}$, i.e., $\mathcal{G}_{t_j} = \sigma(J_{t_1}, \ldots, J_{t_j}) \vee \mathcal{F}_{t_j}$. Noting that $\{(X_{t_j}, Y_{t_j}^c),\ j = 0, 1, \ldots\}$ is jointly normal and is independent of J, we exploit the *innovation*

$$M_{t_{j+1}} = \Delta Y_{t_{j+1}}^c - E\left[\Delta Y_{t_{j+1}}^c \middle| \mathcal{G}_{t_j}\right], \tag{5}$$

which is a sequence of independent normal random variables, to obtain a linear recursive equation that defines $\bar{X}$:

$$\bar{X}_{t_{j+1}} = \lambda_j M_{t_{j+1}} + \beta_j \bar{X}_{t_j} = \lambda_j \Delta Y_{t_{j+1}}^c + \eta_j \bar{X}_{t_j} \tag{6}$$

for suitable coefficients λ_j, β_j, and η_j. We calculate these coefficients in Propositions 2.1 and 2.2 of Section 2.

The second stage consists of filtering the pseudo filter $\bar{X}$ to the original filtration $\mathcal{F}$. In other words,

$$\hat{X}_{t_j} = E\Big[E\big[X_{t_j}\big|\mathcal{G}_{t_j}\big]\Big|\mathcal{F}_{t_j}\Big] = E\big[\bar{X}_{t_j}\big|\mathcal{F}_{t_j}\big] \tag{7}$$

which is valid since $\mathcal{G}_{t_j}$ contains $\mathcal{F}_{t_j}$. Unfortunately, non-linearity appears in this procedure as we mentioned earlier, and knowing $\hat{X}_{t_j}$ is not enough to update $\hat{X}_{t_{j+1}}$. For this reason, we first find the conditional characteristic function $E\Big[e^{i\theta\bar{X}_{t_{j+1}}}\Big|\mathcal{F}_{t_j}\Big]$, which can be updated recursively.

Theorem 1.1. *Let ϕ be the standard normal density, and let $\hat{f}_{j+1}$ be the characteristic function of $M_{t_{j+1}} + \Delta J_{t_{j+1}}$ where $M_{t_{j+1}}$ is defined by (5). Then $H_{j+1}(\theta) = E\Big[e^{i\theta\bar{X}_{t_{j+1}}}\Big|\mathcal{F}_{t_j}\Big]$ satisfies the following recursive equation:*

$$\begin{aligned} &H_{j+1}(\theta) \int \hat{f}_{j+1}(\omega)\, e^{-i\omega\Delta Y_{t_{j+1}}} H_j\big(\gamma_j\omega\big)\, d\omega = \\ &\qquad \sqrt{2\pi}\,\phi(\lambda_j\sigma_{j+1}\theta) \int \hat{f}_{j+1}(\omega)\, e^{-i\omega\Delta Y_{t_{j+1}}} e^{-\omega\lambda_j\sigma_{j+1}^2\theta} H_j\big(\gamma_j\omega + \beta_j\theta\big)\, d\omega \end{aligned} \tag{8}$$

where $\sigma_{j+1} = \mathrm{Var}(M_{t_{j+1}})$ and $\gamma_j = \lambda_j^{-1}(\beta_j - \eta_j)$; λ_j,β_j, and η_j are as in (6).

The proof of the above theorem is given in Section 3. Since $M_{t_{j+1}}$ is a $N(0,\sigma_{j+1}^2)$ random variable independent of $\Delta J_{t_{j+1}}$, one obtains $\hat{f}_{j+1}(\omega)$ as the product of $\exp(-\sigma_{j+1}^2\omega^2/2)$ and the characteristic function of $\Delta J_{t_{j+1}}$. (See Feller (1971), Vol. 2, for the characterization of the infinitely divisible characteristic functions as well as examples.) Differentiating $H_{j+1}(\theta)$ with respect to θ and evaluating at 0, one obtains the optimal filter:

Corollary 1.1. *Let $H_j^\star(\omega) = H_j\big(\gamma_j\omega\big)$ and $H_j^{\star\star}(\omega) = \dfrac{d}{d\theta}H_j(\theta)\Big|_{\theta=(\beta_j-1)\omega}$. Then, $\hat{X}_{t_{j+1}}$ satisfies the following equation:*

$$\begin{aligned} &\hat{X}_{t_{j+1}} \int \hat{f}_{j+1}(\omega)\, e^{-i\omega\Delta Y_{t_{j+1}}} H_j^\star(\omega)\, d\omega = \\ &\qquad i\lambda_j\sigma_{j+1}^2 \int \hat{f}_{j+1}(\omega)\, e^{-i\omega\Delta Y_{t_{j+1}}}\, \omega\, H_j^\star(\omega)\, d\omega - i\beta_j \int \hat{f}_{j+1}(\omega)\, e^{-i\omega\Delta Y_{t_{j+1}}} H_j^{\star\star}(\omega)\, d\omega\,. \end{aligned} \tag{9}$$

(8) and (9) describe the optimal filter recursively. Unfortunately, evaluating these equations may require considerable computation. For each j, one must evaluate H_j at densely sampled θ to update H_{j+1}, and this requires evaluating Fourier integrals as many as the sample size of θ. Consequently this method may be impractical when observations arrive frequently. Because of these problems, we propose an alternative filter that we obtain by approximating $H_j^\star$ and $H_j^{\star\star}$ in (9). It can be shown that $\gamma_j = O(t_{j+1} - t_j)$ as $t_{j+1} - t_j \to 0$, and hence, for small $t_{j+1} - t_j$ we may adopt the following approximations:

$$H_j^\star(\theta) \simeq e^{i\theta\gamma_j\tilde{X}_{t_j}} \qquad \text{and} \qquad H_j^{\star\star}(\theta) \simeq i\tilde{X}_{t_j} e^{i\theta\gamma_j\tilde{X}_{t_j}} \tag{10}$$

where $\tilde{X}_{t_j}$ is the output at time t_j of an approximate filter. Substituting (10) into (9) and applying the Fourier inversion, we obtain another filter $\tilde{X}$ that can be obtained from the following recursion:

$$\tilde{X}_{t_{j+1}} = -\lambda_j \sigma_{j+1}^2 \frac{f'_{j+1}}{f_{j+1}} \Big(\Delta Y_{t_{j+1}} - \gamma_j \tilde{X}_{t_j} \Big) + \beta_j \tilde{X}_{t_j} . \tag{11}$$

It turns out that

$$-\sigma_{j+1}^2 \frac{f'_{j+1}}{f_{j+1}}(y) = E\Big[M_{t_{j+1}} \Big| M_{t_{j+1}} + \Delta J_{t_{j+1}} = y \Big]$$

is the minimum variance unbiased estimate of $M_{t_{j+1}}$ given $M_{t_{j+1}} + \Delta J_{t_{j+1}}$, while $\Delta Y_{t_{j+1}} - \gamma_j \tilde{X}_{t_j} = M_{t_{j+1}} + \Delta J_{t_{j+1}}$. Thus, it is reasonable to expect that the sub-optimal filter $\tilde{X}$ defined by (11) will perform well as long as γ_j is small. In Section 4, we use Monte Carlo simulation to compare the performance of this sub-optimal filter with that of the linear filter proposed in Le Breton and Musiela (1993). The simulation results confirm that $\tilde{X}$ performs far better than the linear filter in the presence of sizable jumps. Moreover the performance of $\tilde{X}$ is robust to misspecified noise structures. This is an attractive property; in practice it is difficult to determine the characteristics of the Gaussian and jump noises.

2 Computation of coefficients

In this section, we prove (6) and as a by-product obtain formula for the coefficients that appear in (6) and in Theorem 1.1, namely: λ, β, γ, η, and σ. Recall that $\mathcal{G}_{t_j}$ is the sigma field generated by $Y^c_{t_1}, \ldots, Y^c_{t_j}$ as well as $J_{t_1}, \ldots, J_{t_j}$, and that the innovation M is a martingale difference sequence defined by

$$M_{t_{j+1}} = \Delta Y^c_{t_{j+1}} - E\Big[\Delta Y^c_{t_{j+1}} \Big| \mathcal{G}_{t_j} \Big] \tag{12}$$

with $M_{t_0} = 0$. Since J is independent of Y^c, $E\Big[\Delta Y^c_{t_{j+1}} \Big| \mathcal{G}_{t_j}\Big] = E\Big[\Delta Y^c_{t_{j+1}} \Big| \Delta Y^c_{t_1}, \ldots, \Delta Y^c_{t_j}\Big]$, and the joint normality implies that $M_{t_{j+1}}$ is independent of $\mathcal{G}_{t_j}$.

Proposition 2.1. *Let $A(t)$ satisfy $dA(t) = a(t)A(t)dt$ with $A(0) = 1$, and define*

$$\beta_j = A(t_{j+1})/A(t_j) \qquad \text{and} \qquad p_{j+1} = A^{-1}(t_{j+1}) \int_{t_j}^{t_{j+1}} c(s)A(s)ds . \tag{13}$$

Then $M_{t_{j+1}} = \Delta Y^c_{t_{j+1}} - p_{j+1}\beta_j \bar{X}_{t_j}$ and $\bar{X}_{t_{j+1}} = \lambda_j M_{t_{j+1}} + \beta_j \bar{X}_{t_j}$ where

$$\lambda_j = \mathrm{Cov}\Big(M_{t_{j+1}}, M_{t_{j+1}} + \int_{t_j}^{t_{j+1}} c(s)b(s)dB_s - \int_{t_j}^{t_{j+1}} d(s)dW_s \Big) \Big/ p_{j+1} \,\mathrm{Var}\,\Big(M_{t_{j+1}} \Big) . \tag{14}$$

In addition, we obtain $\eta_j = \beta_j - \lambda_j p_{j+1} \beta_j$ as well as $\gamma_j = \beta_j p_{j+1}$.

Proof. Because $dA^{-1}(t) = -a(t)A^{-1}(t)dt$ and $dX_t = a(t)X(t)dt + b(t)dB_t$, integration by parts yields

$$A^{-1}(t)X_t = X_0 + \int_0^t A^{-1}(s)b(s)dB_s\,. \tag{15}$$

In the sequel, $H_t := A^{-1}(t)X_t$ has independent increments. Therefore,

$$\Delta Y^c_{t_{j+1}} = \int_{t_j}^{t_{j+1}} c(s)A(s)H_s ds + \int_{t_j}^{t_{j+1}} d(s)dW_s$$

and $E\big[H_s\big|\mathcal{G}_{t_j}\big] = A^{-1}(t_j)\bar{X}_{t_j}$ whenever $s \geq t_j$. From this, we obtain

$$M_{t_{j+1}} = \int_{t_j}^{t_{j+1}} c(s)A(s)H_s ds + \int_{t_j}^{t_{j+1}} d(s)dW_s - p_{j+1}\beta_j\bar{X}_{t_j}\,. \tag{16}$$

On the other hand, integration by parts yields

$$\begin{aligned}\int_{t_j}^{t_{j+1}} c(s)A(s)H_s ds &= H_{t_{j+1}}\int_{t_j}^{t_{j+1}} c(s)A(s)ds - \int_{t_j}^{t_{j+1}} c(s)b(s)dB(s)\\ &= p_{j+1}X_{t_{j+1}} - \int_{t_j}^{t_{j+1}} c(s)b(s)dB(s)\,.\end{aligned}$$

Substituting this into (16), one has

$$X_{t_{j+1}} = \frac{1}{p_{j+1}}\Big[M_{t_{j+1}} + \int_{t_j}^{t_{j+1}} c(s)b(s)dB_s - \int_{t_j}^{t_{j+1}} d(s)dW_s\Big] + \beta_j\bar{X}_{t_j}\,. \tag{17}$$

Note that the first term in (17) is independent of $\mathcal{G}_{t_j}$ while $\beta_j\bar{X}_{t_j}$ is measurable with respect to $\mathcal{G}_{t_j}$. Thus,

$$\hat{X}_{t_{j+1}} = \frac{1}{p_{j+1}}E\Big[M_{t_{j+1}} + \int_{t_j}^{t_{j+1}} c(s)b(s)dB_s - \int_{t_j}^{t_{j+1}} d(s)dW_s\Big|M_{t_{j+1}}\Big] + \beta_j\bar{X}_{t_j}$$

and the rest follows from the joint normality.

□

In the following proposition, we find a recursive expression for (14).

Proposition 2.2. *Let $\sigma^2_{j+1} = \mathrm{Var}(M_{t_{j+1}})$ and $v_{j+1} = E\big[(X_{t_{j+1}} - \bar{X}_{t_{j+1}})^2\big]$. Then*

$$\lambda_j = \frac{\sigma^2_{j+1} + r_{j+1}}{p_{j+1}\sigma^2_{j+1}} \tag{18}$$

$$\sigma^2_{j+1} = \int_{t_j}^{t_{j+1}} \Big(b^2(s)\Big[p_{j+1}\frac{A(t_{j+1})}{A(s)} - c(s)\Big]^2 + d^2(s)\Big)\,ds + (p_{j+1}\beta_j)^2 v_j \tag{19}$$

$$v_{j+1} = p_{j+1}^{-2}\int_{t_j}^{t_{j+1}} \Big(c^2(s)b^2(s) + d^2(s)\Big)\,ds - p_{j+1}^{-2}\sigma^2_{j+1}(1 - p_{j+1}\lambda_j)^2 \tag{20}$$

where

$$r_{j+1} = \int_{t_j}^{t_{j+1}} \Big(b^2(s)c(s)\Big[p_{j+1}\frac{A(t_{j+1})}{A(s)} - c(s)\Big] - d^2(s)\Big)\,ds\,.$$

Proof. Note that

$$\begin{aligned}
\int_{t_j}^{t_{j+1}} c(s)A(s)H_s ds &= \int_{t_j}^{t_{j+1}} c(s)A(s)(H_s - H_{t_j})ds + p_{j+1}\beta_j X_{t_j} \\
&= (H_{t_{j+1}} - H_{t_j})\int_{t_j}^{t_{j+1}} c(s)A(s)ds - \int_{t_j}^{t_{j+1}} c(s)b(s)dB_s + p_{j+1}\beta_j X_{t_j}
\end{aligned}$$

where the last equality is obtained via integration by parts with $dH_s = b(s)A^{-1}(s)dB_s$. Then (16) can be rewritten as

$$M_{t_{j+1}} = \int_{t_j}^{t_{j+1}} \Big[p_{j+1}\frac{A(t_{j+1})}{A(s)} - c(s)\Big] b(s)dB_s + \int_{t_j}^{t_{j+1}} d(s)dW_s + p_{j+1}\beta_j(X_{t_j} - \bar{X}_{t_j}). \quad (21)$$

Therefore (19) follows immediately. (21) also implies that

$$\mathrm{Cov}\Big(M_{t_{j+1}}, \int_{t_j}^{t_{j+1}} c(s)b(s)dB_s - \int_{t_j}^{t_{j+1}} d(s)dW_s\Big) = r_{j+1}$$

and hence (18) follows. Next, subtracting $\bar{X}_{t_{j+1}} = \lambda_j M_{t_{j+1}} - \beta_j \bar{X}_{t_j}$ from (17), one has

$$p_{j+1}(X_{t_{j+1}} - \bar{X}_{t_{j+1}}) = (1 - p_{j+1}\lambda_j)M_{t_{j+1}} + \int_{t_j}^{t_{j+1}} c(s)b(s)dB_s - \int_{t_j}^{t_{j+1}} d(s)dW_s.$$

On the other hand, (14) implies that

$$\mathrm{Cov}\Big(M_{t_{j+1}}, \int_{t_j}^{t_{j+1}} c(s)b(s)dB_s - \int_{t_j}^{t_{j+1}} d(s)dW_s\Big) = -\sigma_{j+1}^2(1 - p_{j+1}\lambda_j).$$

Therefore

$$p_{j+1}^2 v_{j+1} = \int_{t_j}^{t_{j+1}} \Big(c^2(s)b^2(s) + d^2(s)\Big)\, ds - \sigma_{j+1}^2(1 - p_{j+1}\lambda_j)^2$$

and we obtain (20).

□

3 Proof of Theorem 1.1

We will prove Theorem 1.1 by evaluating $H_{j+1}(\theta) = E\Big[e^{i\theta \bar{X}_{t_{j+1}}}\Big|\mathcal{F}_{t_{j+1}}\Big]$ for a fixed but an arbitrary j. Both $\bar{X}_{t_{j+1}}$ and $\Delta Y_{t_{j+1}}$ depend on $\mathcal{F}_{t_j}$, and this dependence makes it difficult to compute the conditional expectation. In Gaussian cases, one can use an innovation process to remove such dependencies. However such methods generally do not exist for non-Gaussian noise. Instead, we find another probability measure which makes $(\Delta Y^c_{t_{j+1}}, \Delta J_{t_{j+1}})$ independent of $\mathcal{G}_{t_j}$ (and hence independent of $\mathcal{F}_{t_j}$). Throughout this section, P donotes the probability measure associated with our model. The following lemma is parallel to Girsanov's theorem:

Lemma 3.1. *Define a probability measure $Q = Q_{j+1}$ via*

$$\frac{dP}{dQ} = L_{t_{j+1}} = \exp\Big(\frac{\gamma_j \bar{X}_{t_j}}{\sigma_{j+1}^2} \Delta Y_{t_{j+1}}^c - \frac{1}{2} \frac{\gamma_j^2 \bar{X}_{t_j}^2}{\sigma_{j+1}^2} \Big). \tag{22}$$

Then the law of $(M_{t_1}, \ldots, M_{t_j}, \Delta Y_{t_{j+1}}^c)$ under Q is the same as the law of $(M_{t_1}, \ldots, M_{t_{j+1}})$ under P. In particular, $\Delta Y_{t_{j+1}}^c$ is independent of $\mathcal{G}_{t_j}$.

Proof. Note that

$$L_{t_{j+1}}^{-1} = \exp\Big(-\frac{\gamma_j \bar{X}_{t_j}}{\sigma_{j+1}^2} M_{t_{j+1}} - \frac{1}{2} \frac{\gamma_j^2 \bar{X}_{t_j}^2}{\sigma_{j+1}^2} \Big)$$

is a positive random variable of mean 1 with respect to P, and hence Q is well defined. Also note that, under P, $M_{t_{j+1}} = \Delta Y_{t_{j+1}}^c - \gamma_j \bar{X}_{t_j}$ is a $N(0, \sigma_{j+1}^2)$ random variable independent of $\mathcal{G}_{t_j}$. Therefore

$$E_P\Big[e^{i\theta \Delta Y_{t_{j+1}}^c} L_{t_{j+1}}^{-1} \,\Big|\, \mathcal{G}_{t_j} \Big] = e^{-\theta^2 \sigma_{j+1}^2 / 2} \tag{23}$$

almost surely, and the result follows.

□

In what follows, E_Q denotes the expectation under Q, while E_P, or simply E, is used for the expectation under P. We will use the following version of Bayes' rule:

$$H_{j+1}(\theta) \cdot E_Q\Big[L_{t_{j+1}} \,\Big|\, \mathcal{F}_{t_{j+1}} \Big] = E_Q\Big[e^{i\theta \bar{X}_{t_{j+1}}} L_{t_{j+1}} \,\Big|\, \mathcal{F}_{t_{j+1}} \Big]. \tag{24}$$

(See Karatzas and Shreve (1987), p193.) The next step is to prove the following:

$$\begin{aligned} &E_Q\Big[e^{i\theta \bar{X}_{t_{j+1}}} L_{t_{j+1}} \,\Big|\, \mathcal{F}_{t_{j+1}} \Big] \cdot f_{j+1}(\Delta Y_{t_{j+1}}) = \\ &\frac{1}{\sqrt{2\pi}} \phi(\lambda_j \sigma_{j+1} \theta) \int \hat{f}_{j+1}(\omega)\, e^{-i\omega \Delta Y_{t_{j+1}}} e^{-\omega \lambda_j \sigma_{j+1}^2 \theta} H_j\Big(\gamma_j \omega + \beta_j \theta \Big)\, d\omega . \end{aligned} \tag{25}$$

Then $E_Q\Big[L_{t_{j+1}} \,\Big|\, \mathcal{F}_{t_{j+1}} \Big] f_{j+1}(\Delta Y_{t_{j+1}})$ can be obtained by evaluating (25) at $\theta = 0$, and hence (25) completes the proof of Theorem 1.1.

Lemma 3.2. *Let $Y = W + Z$, where W is a $N(0, \sigma^2)$ random variable independent of Z. Then*

$$f(y) \cdot E\Big[e^{i\xi W} \exp\Big(\frac{x}{\sigma^2} W - \frac{1}{2} \frac{x^2}{\sigma^2} \Big) \,\Big|\, Y = y \Big] = \frac{e^{i\xi x}}{\sqrt{2\pi}} \phi(\xi \sigma) \int \hat{f}(w) e^{-\xi \sigma^2 \omega} e^{-i(y-x)\omega}\, d\omega \tag{26}$$

where f is the density of Y, $\hat{f}$ is the corresponding characteristic function, and ϕ is the standard normal density.

Proof. Note that the Fourier transform of $f(y)\, E\Big[e^{i\xi W} \exp\Big(\frac{x}{\sigma^2}W - \frac{1}{2}\frac{x^2}{\sigma^2}\Big) \,\Big|\, Y = y\Big]$ is identical to $E\Big[e^{i\omega Y} e^{i\xi W} \exp\Big(\frac{x}{\sigma^2}W - \frac{1}{2}\frac{x^2}{\sigma^2}\Big)\Big]$, which further reduces to

$$E\Big[e^{i\omega Z}\Big] \cdot E\Big[e^{i(\omega+\xi)W} \exp\Big(\frac{x}{\sigma^2}W - \frac{1}{2}\frac{x^2}{\sigma^2}\Big)\Big] = \hat{f}(\omega)\, e^{\omega^2\sigma^2/2} \cdot e^{i(\omega+\xi)x - \frac{1}{2}(\omega+\xi)^2\sigma^2}. \tag{27}$$

The inverse Fourier transform of (27) is identical to the right side of (26), and the result follows.

□

Now we resume the proof of Theorem 1.1. Recall that $\bar{X}_{t_{j+1}} = \lambda_j \Delta Y^c_{t_{j+1}} + \eta_j \bar{X}_{t_j}$, where $\eta_j = \beta_j - \lambda_j \gamma_j$. Since $\bar{X}_{t_j}$ is $\mathcal{G}_{t_j}$-measurable, we obtain

$$E_Q\Big[e^{i\theta \bar{X}_{t_{j+1}}} L_{t_{j+1}} \,\Big|\, \mathcal{F}_{t_{j+1}}\Big] = E_Q\Big[e^{i\theta\eta_j \bar{X}_{t_j}}\, E_Q\Big(e^{i\theta\lambda_j \Delta Y^c_{t_{j+1}}} L_{t_{j+1}} \,\Big|\, \mathcal{G}_{t_j},\, \Delta Y_{t_{j+1}}\Big) \,\Big|\, \mathcal{F}_{t_{j+1}}\Big]. \tag{28}$$

Under Q, $\Delta Y^c_{t_{j+1}}$ is a $N(0, \sigma^2_{j+1})$ random variable independent of $\Delta J_{t_{j+1}}$ and f_{j+1} is the density of $\Delta Y_{t_{j+1}}$. Therefore, (26) and (28) yield

$$\begin{aligned} &E_Q\Big[e^{i\theta \bar{X}_{t_{j+1}}} L_{t_{j+1}} \,\Big|\, \mathcal{G}_{t_j},\, \Delta Y_{t_{j+1}}\Big] \cdot f_{j+1}(\Delta Y_{t_{j+1}}) = \\ &\quad \frac{1}{\sqrt{2\pi}}\, \phi(\theta\lambda_j\sigma_{j+1}) \int \hat{f}_{j+1}(\omega)\, e^{-i\omega \Delta Y_{t_{j+1}}} e^{-\omega\lambda_j\sigma^2_{j+1}\theta} e^{i(\gamma_j\omega + \beta_j\theta)\bar{X}_{t_j}}\, d\omega\,. \end{aligned} \tag{29}$$

The last step is to compute the conditional expectation of (29) with respect to $\mathcal{F}_{t_{j+1}}$. Note that $\Delta Y_{t_{j+1}}$ is $\mathcal{F}_{t_{j+1}}$-measurable and

$$E_Q\Big[e^{i(\gamma_j\omega+\beta_j\theta)\bar{X}_{t_j}} \,\Big|\, \mathcal{F}_{t_{j+1}}\Big] = E_P\Big[e^{i(\gamma_j\omega+\beta_j\theta)\bar{X}_{t_j}} \,\Big|\, \mathcal{F}_{t_j}\Big] = H_j\Big(\gamma_j\omega + \beta_j\theta\Big)\,.$$

(This last is a consequence of the above lemma.) Therefore the conditional expectation of (29) with respect to $\mathcal{F}_{t_{j+1}}$ yields (25), and the proof is complete.

4 Numerical Results

Because the optimal filter $\hat{X}$ requires excessive computation, we proposed the sub-optimal filter $\tilde{X}$ defined in (11). Provided that observations arrive frequently, the performance of $\tilde{X}$ is expected to be near by optimal. In this section, we show simulation results in which $\tilde{X}$ is compared with the best linear filter obtained by Le Breton and Musiela (1993). We also provide simulation results that show how $\tilde{X}$ performs when the correct recognition of the noise structure fails.

In this section, we consider observations that arrive at a fixed rate, or equivalently the arrival time of the j-th observation is $j\delta$, and the last arrival time is denoted by T. The coefficients of the signal process (1) and the observation process (2), namely a, b, c, and d,

are set to be identically 1, and the signal at time 0 is also 1. In this case, it turns out that $\beta_j = e^{\delta}$, $\gamma_j = \beta_j - 1$, and

$$\begin{aligned} \sigma_{j+1} &= \frac{(\beta_j-1)^2}{2\beta_j^2}(\beta_j^2-1) - 2\frac{(\beta_j-1)^2}{\beta_j} + \delta + (\beta_j-1)^2 v_j\,, \\ \lambda_j &= \frac{\beta_j}{\beta_j-1} + \sigma_{j+1}^{-2}\Big[\beta_j - 1 - 2\beta_j(\beta_j-1)^{-1}\delta\Big]\,, \\ v_{j+1} &= \beta_j^2(\beta_j-1)^{-2}(2\delta - \beta_j^2\sigma_{j+1}^2)\,, \end{aligned}$$

while $v_0 = 0$. Thus one updates σ_{j+1}, λ_j, and v_{j+1} in turns. We take J to be a symmetric alpha stable process. Such processes have been used for modeling heavy tailed noise; see, e.g., Nikias and Shao (1995). More precisely, the characteristic function of J_t is given by

$$E\Big[e^{i\theta J_t}\Big] = \exp(-t\zeta|\theta|^{\alpha})$$

where α is a stable index and ζ is a dispersion parameter. Different values of these parameters will be used depending upon the purpose of simulations.

To implement $\tilde{X}$ one must evaluate $\sigma_{j+1}^2 f'_{j+1}/f_{j+1}$, but no closed form expression for f_{j+1} is available. One can, however, use Fourier inversion, as the Fourier transform $\hat{f}_{t_j}$ of f_{j+1} is well known. In order to reduce round-off errors, we use the standard scale: there is an appropriate constant $\varepsilon_{j+1} > 0$ such that the law of $(M_{t_{j+1}} + \Delta J_{t_{j+1}})/\sigma_{j+1}$ is the same as that of $M + \varepsilon_{j+1}J$ where M is standard normal and J is symmetric alpha stable with dispersion 1. Consequently,

$$\sigma_{j+1}^2 \frac{f'_{j+1}}{f_{j+1}}(y) = \sigma_{j+1}\frac{h'_{j+1}}{h_{j+1}}\left(\frac{y}{\sigma_{j+1}}\right)$$

where h_{j+1} is the density of $M + \varepsilon_{j+1}J$. The fast Fourier transform is used for evaluating h_{j+1}, the inverse Fourier transform of $\exp(-\theta^2/2 - |\varepsilon_{j+1}\theta|^{\alpha})$.

4.1 Comparison between $\tilde{X}$ and Linear Filter

The performance of $\tilde{X}$ is compared with that of the best linear filter given by Le Breton and Musiela (1993). Three different stable indices (1.0, 1.4, and 1.8) were taken, while the dispersion parameter is set to be 1. The inter-arrival time of observations is 0.01 and the expiry $T = 10$. Le Breton and Musiela's filter minimizes L^p error (for a specified value of $p > 1$) among all linear filters. Thus it is not applicable when the stable index is 1.0. $p = 1.1$ was used in the other cases.

100,000 Monte Carlo simulations were generated to estimate L^p error and L^2 error, and the results are summarized in Table 1. Our filter is denoted by 'AF' while 'LM' is used for Le Breton and Musiela's filter. For comparison, we also implemented the pseudo filter ($\bar{X}$) using the same simulation, except that the non-Gaussian part of the observation noise was omitted. The L^2 error in this case is 2.414, and so is not substantially better than that

of $\tilde{X}$. (In fact, the L^2 error of the psuedo filter sets a lower bound of the L^2 errors of any filter.) Since the average value of X_T is 20959, the relative L^2 error of $\tilde{X}$ is quite small. Also $\tilde{X}$ substantially outperforms the best linear filter in both L^2 and L^p errors. As the sample variance of alpha stable distribution diverges in the order of $n^{2/\alpha-1}$ where n denotes the sample size, the L^2 error of the linear filter is decreasing in α as shown in the table.

	L^p error		L^2 error	
	AF	LM_p	AF	LM_p
$\alpha = 1.0$	1.75636	n/a	4.14959	n/a
$\alpha = 1.4$	1.68752	4.59625	3.86310	180.25826
$\alpha = 1.8$	2.11866	4.17176	5.81958	28.17528

Table 1: Performances of $\tilde{X}$ (AF) and Le Breton-Musiela $p = 1.1$ (LM_p)

4.2 Robustness

Ideally, filter design proceeds from complete a priori knowledge of the structure of the observation noise. The required parameters can be estimated from the history. It is important to check the sensitivity of the filter performance to departure of the assumed value of α from the true value.

	AF $\alpha = 1.0$	AF $\alpha =$ true	Kalman
$\alpha = 1.0$	3.09941	3.09941	15903.76435
$\alpha = 1.4$	3.21928	3.07577	20.25297
$\alpha = 1.7$	3.19790	3.00424	3.02967
$\alpha = 1.8$	3.19219	2.99719	2.63541

Table 2: Robustness of $\tilde{X}$ (AF $\alpha = 1.0$)

Data in Table 2 show that $\tilde{X}$ is robust with respect to uncertainty about the value of the stable index. In order to obtain a fair comparison, we do not vary the dispersion parameter which we set 0.1. Observations arrive each 0.01 time unit and the expiry $T = 5$. As before, 100,000 simulation runs were produced to estimate the L^2 error. We focus on the performance of *Cauchy* filter: that is, $\tilde{X}$ was designed for the stable index 1.0. The first column shows how this Cauchy filter performs when the actual stable index is different

from 1.0. Four different values (1.0, 1.4, 1.7, and 1.8) of stable index were considered, and the performance was robust to this misspecification. The second column shows the performance of $\tilde{X}$ when the stable index α is correctly specified. As shown in the table, the degradation due to use of the wrong value of α was not substantial. The last column shows the performance of Kalman filter. The average of X_T was 148.4, and the L^2 error for the pseudo filter was 2.414. Thus we may conclude that the Cauchy filter estimates the signal accurately as well. When the stable index is near 2.0, it turns out that the Kalman filter slightly outperforms $\tilde{X}$. This is due to the small dispersion parameter, as well as large stable index, which makes the observation noise virtually Gaussian.

5 Summary and Remarks

In order to extract a Gaussian signal contaminated by an infinitely divisible noise, such as compound Poisson or alpha stable, we constructed the filter $\hat{X}$ that minimizes the L^2 error. Since implementing this optimal filter requires excessive computation, we propose a more practical filter $\tilde{X}$ that approximates the optimal filter. In fact, if the function $\sigma_{j+1}f'_{j+1}/f_{j+1}$ defined in (11) is Lipschitz uniformly as $\max_j |t_{j+1} - t_j| \to 0$, one can show that $\tilde{X} - \hat{X}$ converges to 0 in L^2 uniformly on any compact interval. So far, we have not been able to check whether a specific model satisfies this assumption. In this paper we include simulation results which show that $\tilde{X}$ outperforms the existing best linear filter, provided that observations arrive sufficiently frequently. We also show that the performance is insensitive to the misspecification of the observation noise distribution. Although the robustness of $\tilde{X}$ enables us to obtain an acceptable estimate of the signal even with an inaccurate description of the observation noise, a sound statistical procedure for identifying the infinitely divisible noise would still be of considerable benefit and should be investigated in the future.

A significant limitation of the results in this paper is that the signal is restricted to be Gaussian. In many applications of signal detection, it is inappropriate to describe a signal as a Gaussian process. Although it is questionable whether the L^2 optimal filter can be obtained as a recursive algorithm, a tractable sub-optimal filter may be available and could be useful.

References

Cambanis, S. and G. Miller, (1981) Linear problems in pth order and stable processes. *SIAM Appl. Math.* **41** 43-69.

Devroye, L., (1986) Non-Uniform Random Variate Generation. (Springer-Verlag, New York).

Feller, W., (1971) An Introduction to Probability Theory and its Application, Vol. 2, (John Wiley, New York).

Karatzas, I. and S. Shreve, (1987) Brownian Motion and Stochastic Calculus. (Springer-Verlag, New York).

Kassam, S. A., (1988) Signal Detection in Non-Gaussian Noise (Springer-Verlag, New York).

Le Breton, A. and M. Musiela, (1993) A generalization of the Kalman filter to models with infinite variance. *Stoch. Proc. and their Appl.* **47** 75-94.

Nikias, c. L. and M. Shao, (1995) Signal Processing with Alpha-Stable Distributions and Applications. (Wiley, New York).

Poor, H. V., (1988) Fine quantization in signal detection and estimation. *IEEE Info* **34** 960-972.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, (1992) Numerical Recipes in C: the art of scientific computing. (Cambridge University Press, Cambridge).

Protter, P., (1990), Stochastic Integration and Differential Equations. (Springer-Verlag, New York).

Samorodnitsky, G. and M. S. Taqqu, (1994) Stable Non-Gaussian Random Processes. (Chapman & Hall, New York).

Stuck, B. W., (1978) Minimum error dispersion linear filtering of scalar symmetric processes, *IEEE Trans. Aut. Control* AC-23 507-509.